

인터넷 오디오 방송을 위한 오디오 스트리밍 기술

강경욱, 홍진우

한국전자통신연구원

Audio Streaming Technology for Internet Audio Broadcasting

Kyeongok Kang, Jin-Woo Hong

ETRI

kokang@etri.re.kr, jwhong@etri.re.kr

요약

인터넷을 이용한 디지털 오디오 방송 서비스에 대한 관심이 집중되면서 디지털 오디오 데이터를 서버로부터 사용자에게 까지 실시간으로 전송하기 위한 연구가 진행되고 있다. 디지털 오디오 방송의 실시간 전송을 위하여 효율적인 오디오 압축 기술의 개발도 중요하지만 이들 오디오 압축 기술과 연계되는 오디오 스트리밍 기술이 매우 중요하다.

본 논문에서는 현재 사용되고 인터넷 오디오 방송 관련 기술을 분석하고, 특히, IETF에서 논의되고 있는 MPEG-2 AAC 및 MPEG-4 오디오를 인터넷을 통하여 전송하기 위한 RTP payload 포맷을 분석하고, 기술개발을 위한 고려사항을 제안한다.

1. 서론

라디오도 디지털 시대로 들어감에 따라 많은 라디오 방송국도 현재 변화의 물결에 직면하고 있으며, 머지 않은 시간에 디지털 오디오 방송(DAB)은 라디오의 분배 측면에 멀티미디어 성분을 추가할 것이다. 뿐만 아니라 더 이상 전적으로 디스크나 지상/케이블/위성 채널만을 통한 오디오의 수신은 없을 것이며, IP(Internet Protocol)를 통한 오디오 정보의 전송, 오디오의 다운로드 및 온라인 스트리밍(streaming) 등과 같은 서비스가 활성화 될 것으로 예측되고 있다.

방송사업자들에 있어서 인터넷은 초기의 방송 정보를 올리는 사이트라는 위치에서, 웹에서의 push 및 pull 기술을 이용한 오디오 스트리밍과 양방향(interactive)의 개념을 통하여 새로운 고객에 접근할 수 있는 가치 창출의 부가적인 미디어로 그 역할이 변화되고 있다. 증가된 대역폭의 유용성, 더 빠른 모뎀의 사용, 더욱 개선되고 scalability 기능의 오디오 부호화 방법의 개발 등으로 인하여 웹에서 더욱 향상된 품질의 오디오를 들을 수 있는 기회가 증가될 것이다. 그러나, 현재 웹은 여전히 좁은 대역폭 밖에 제공하지 못하고 있으며, 이는 오디오 신호를 인터넷을 통해 전송하기 전에 상당히 복잡한 압축 기술이 오디오 신호에 적용되어야 함을 의미한다.

인터넷 라디오는 생방송이든 요구형(on-demand)이든 스트리밍 오디오 기술을 필요로 한다. 이를 위해서는 수신단에서의 재생 소프트웨어 뿐만 아니라 송신단

에서는 적절한 전송 프로토콜과 서버 소프트웨어를 포함한 특정 스트리밍 소프트웨어를 필요로 한다. 오디오 신호의 다운로드와 비교해볼 때, 스트리밍 오디오는 오디오 정보의 실시간 전송 및 수신에 관점으로 인하여 더욱 더 복잡한 문제이다.

본 논문에서는 현재 사용되고 인터넷 오디오 방송 관련 기술을 분석하고, 특히, IETF(Internet Engineering Task Force)에서 논의되고 있는 MPEG-2 AAC 및 MPEG-4 오디오를 인터넷을 통하여 실시간으로 전송하기 위한 RTP(Realtime Transport Protocol) payload 포맷을 분석하고, 기술개발을 위한 고려사항을 제안한다.

2. 인터넷 오디오 방송을 위한 스트리밍 기술

2.1 압축 기술의 품질

1997년에 유럽방송연맹(EBU)에서는 인터넷을 통한 오디오 압축 기술에 대한 청취 시험을 수행하였으며, 그 결과를 그림 1에 나타낸다[1]. 그 결과에 의하면, 14.4kbps 아날로그 모뎀을 통한 오디오 품질은 바로 "very annoying"이었다. 28.8kbps 모뎀의 경우에는 오디오 품질은 상당히 개선되었으나, 압축 기술 사이의 차이점은 상당하다. 64kbps의 전송 용량을 가진 B 채널을 사용한 ISDN 라인을 이용한 시험 결과에 의하면, MPEG-Audio Layer II 및 III의 품질 레벨은 모노 신호의 경우에는 좋은 FM 라디오 품질과 비교할 수 있으며, 스테레오 신호의 경우에도 AM 라디오보다 좋다.

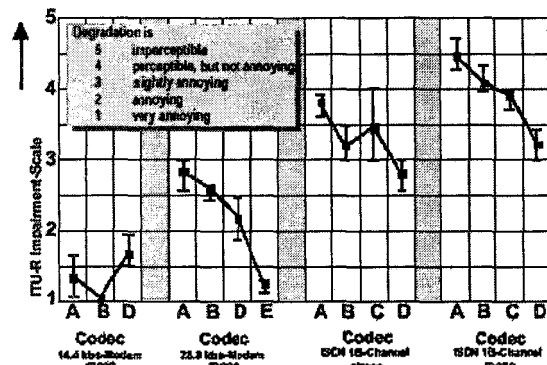


그림 1. 인터넷 오디오 시스템의 품질평가 결과(EBU) (Codec A: MP3, Codec B & C: MPEG-Audio Layer II, Codec D: RealAudio 4.0, Codec E: Eurosoft)

또한, MPEG-2 AAC, Quicktime 의 Musiccodec 등과 같은 새로운 코덱을 시험하기 위한 추가적인 테스트가 독일 IRT 에서 1998 년에 수행되었다. 이때 사용한 테스트 절차는 1997 년의 시험절차와 동일하다. 기준 시스템으로 MPEG-2 와 'MPEG-2.5' Layer III(MP3 로 통용)가 테스트에 포함되었다[1].

그림 2 에 전체 테스트 결과를 나타낸다. 14.4kbps 의 모뎀 속도에서의 테스트 결과는 부정적이었다. 즉, AAC 와 같은 최상의 시스템도 MPEG-2 Layer II 나 MPEG-3 Layer III 에 비해 개선된 결과를 보여주질 못하고 있다. 이와 유사한 결과가 ISDN 모뎀 채널 접속의 경우에도 나타났다. 여기에는 2 가지의 이유가 있다. 첫 번째 이유는 ISDN 모뎀 접속에서 MP3 에 의한 오디오 품질은 이미 상당히 좋은 수준이기 때문이고, 두 번째는 테스트 방법론으로 설명할 수 있다. 주관 평가에 적용된 테스트 방법은 고품질, 즉 거의 투명한(near-transparent) 품질에 초점을 맞춘 것이 아니고 중간(intermediate) 품질에 초점을 맞추었다. 이는 이러한 방법론은 고품질 레벨에서의 작은 차이를 구별할 수 없음을 의미한다.

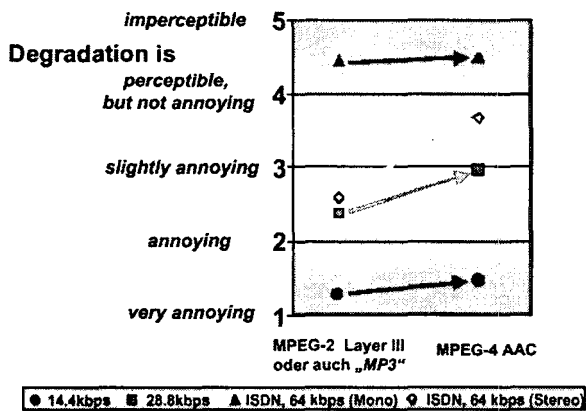


그림 2. 오디오 시스템에 대한 추가실험(1996→1998)

그러나, 새로운 테스트 결과에 의하면, 28.8kbps 에서 모뎀 채널 신호와 ISDN B-channel 접속에서 스테레오 신호에 대해서는 AAC 가 MP3 에 비해 상당히 품질이 개선된 것을 알 수 있다. 적어도 이미 28.8kbps 의 모뎀 속도로 인터넷에 접속하기 때문에 이 결과는 새로운 테스트의 상당히 중요한 결과이다. 또한, 이러한 모뎀으로 스트리밍을 할 수 있는 비트율은 대부분의 경우에 유용하기 때문에 더욱 그러하다. 그림 2 에서 알 수 있는 또 다른 중대한 개선 상황은 스테레오 프로그램을 ISDN 1-B 채널 접속을 통하여 스트리밍할 수 있다는 것이다. ITU-R 의 5 점 손실 척도 등급으로 볼 때 이 경우에 AAC 가 MP3 에 비해 1 등급 정도 향상된 음질을 제공하고 있다.

2.2 오디오 스트리밍 시스템

모든 오디오 스트리밍 기술은 방대한 양의 데이터를 직접 사용하는 대신에 컴팩트한 오디오 파일을 만드는 오디오 신호의 압축을 위한 인코더를 사용한다. CD 오디오를 스트리밍하기 위해서는 1.5Mbps 의 대역폭이 요구되기 때문에 높은 데이터 reduction 은 필요하다.

28.8kbps 의 모뎀의 속도는 CD 에 비해 50 배 정도 느리기 때문에 오디오 인코더에서 약 50 배 정도 압축하여야 한다. 더구나, 아날로그 모뎀 라인은 잡음이나 다른 왜곡으로 이론적인 성능을 결코 얻을 수 없게 때문에, 28.8kbps 모뎀에 요구되는 50 배(50:1) 압축보다 더 압축할 수 있는 인코더를 사용하여야 한다.

ISDN 에 요구되는 압축률은 하나의 B 채널만을 사용하면 24:1 이고, 하나의 ISDN 라인에 의해 제공되는 2 개의 B 채널을 모두 사용하면 12:1 이다. 물론, 인터넷 라디오 청취자가 ISDN 으로 접속하여 28.8kbps 의 파일에 접속할 수 있다. 이 경우는 빠른 대역폭 접속을 사용하였으나 ISDN 라인이 가지는 장점을 전혀 받지 못하게 된다.

오디오 파일의 다운로드 또는 라이브 오디오 스트리밍에 사용되고 있는 시스템의 개요를 소개한다[1].

- **AAC(Advanced Audio Coding)** : MPEG-2 AAC 및 추가 기능과 함께 MPEG-4 오디오 표준의 하나로 사용되고 있으며, 현재 인터넷 응용으로 유용한 것은 Liquid Audio 에서 구현한 것이 유일하다. Liquid Audio 에서 구현한 AAC 의 약점은 일반적으로 라이브 스트리밍 또는 통상적인 서버로부터 AAC 부호화 파일을 재생하는 것을 허용하지 않는다는 것이다. 이 시스템은 현재 인터넷 상에서의 music distribution 에만 적용된다.
- **Microsoft Media 4 Audio** : WMT(Windows Media technologies) 4.0 에 기반을 두고 있다. 음악과 음성 콘텐츠를 부호화할 수 있도록 특별히 고안된 2 개의 코덱을 가지고 있다. 인코딩 속도는 오히려 빨라 표준 PC 에서 실시간 인코딩이 가능하며 RealNetworks G2 에 비교될 수 있다. MS Media 4 Audio 는 모뎀 및 스테레오 신호에 대해 8kHz 에서 48kHz 까지의 표본화 주파수로 5 ~ 128kbps 의 광범위한 비트율 영역을 제공한다. 음성의 부호화를 위하여 매우 낮은 비트율에서 고품질의 광대역 신호를 재생할 수 있는 음성 코덱을 사용한다. WMT 4.0 을 사용하여, 콘텐츠 제공자는 요구형(On-demand) 및 라이브 스트림 모두에 대해 5 개 까지의 서로 다른 비트율(multi-bit-rate) 의 스트림을 하나의 ASF(Advanced Streaming Format) 파일의 형태로 제공할 수 있다.
- **MP3** : 오디오 파일의 스트리밍 또는 다운로드에 주로 사용하는 특정 파일 포맷을 규정하고 있으며 최근에는 라이브 스트리밍도 가능하다. MP3 는 프라운호퍼 연구소에서 특허권을 가지고 있는 시스템이며 ISO/IEC MPEG Layer III 에 기반으로 두고 있으나, 저 비트율에서는 이 표준을 따르지 않는 등 항상 이 표준을 따르는 것은 아니다. 현재 시장에는 여러 가지 버전의 MP3 가 존재한다.
- **Q-Design Music Codec 2** : Quicktime 4.0 멀티미디어 플랫폼에서 동작하는 이 시스템은 오디오 및/또는 비디오의 다운로드 뿐만 아니라 1999 년 4 월에 Quicktime 4.0 베타 버전을 발표한 이후 라이브 스트리밍도 지원하고 있다. 버전 3에서는 오직 오디오 파일의 스트리밍만 가능하였으나, 1999 년 6 월 이후에는 완전한 버전의 Quicktime 4.0 이 유용하게 되었다. 완전히 새로운 특허권을 가진 파라메트릭 부호화 시스템에 기반을 둔 Music Codec 2 는 매우 낮은

비트율에서도 상당히 좋은 품질을 제공한다.

- **RealAudio G2 system** : 오디오의 라이브 스트리밍이나 오디오 파일의 스트리밍에 사용되고 있다. 새로운 G2 시스템은 DolbyNet 부호화 기술에 기반을 둔 RealAudio 5.0에 비해 상당히 진보되었다. G2 시스템은 더 좋은 품질 뿐만 아니라 scalability도 제공한다. 이것은 14.4kbps의 모뎀 뿐만 아니라 64kbps의 ISDN에서 동시에 사용될 수 있다. 최대 6개까지의 많은 병렬 스트림이 하나의 오디오 파일내에 동시에 생성될 수 있다. 이 시스템은 인터넷이 병목 현상에 직면했을 경우에는 품질을 떨어뜨려 조절한다.

3. MPEG 오디오 스트리밍 기술

다양한 오디오 및 비디오 압축기술을 실시간 인터넷 서비스에 응용하기 위하여 IETF(Internet Engineering Task Force) 산하의 AVT WG(Audio/Video Transport Working Group)에서는 MPEG-1/2 Audio, MPEG-2 AAC(Advanced Audio Coding) 및 MPEG-4 Audio 데이터 전송에 적합한 RTP(Realtime Transport Protocol) payload 포맷을 제안하기 위한 작업을 수행하고 있다.

본 장에서는 MPEG-2 AAC 및 MPEG-4 Audio에 대한 RTP payload 포맷을 간략히 기술한다.

3.1 MPEG-2 AAC

ISO/IEC MPEG-2 AAC(Advanced Audio Coding) 기술은 채널 당 64kbps 이하의 전송률에서 탁월한 오디오 품질을 제공한다[2]. 이는 1개에서 48개까지의 오디오 채널, 16개까지의 저 주파수 대역(subwoofer) 채널, 16개까지의 부가(embedded) 데이터 채널을 지원하는 유연한 비트스트림 신택스를 가지고 있다. AAC는 넓은 영역의 표본화 주파수(16kHz ~ 96kHz)를 지원하여 상당히 광범위한 영역의 비트 율을 제공할 수 있다. 따라서, 프로급 또는 가정에서의 극장 사운드 시스템에서부터 인터넷 음악 방송 시스템에 이르는 영역까지의 넓은 응용 범위를 지원한다.

AAC는 높은 신호 압축을 얻기 위하여 고 분해능의 필터 뱅크, 강력한 오디오 자각 모델, 역방향 적응 예측, joint channel 부호화 및 Huffman 부호화 등의 효율적인 부호화 방법을 결합한 것이다. 1998년에 MPEG 오디오 서브그룹에서 MPEG 오디오 부호화기 계열에 대한 테스트를 수행한 결과에 의하면, 스테레오 신호에 대하여 96kbps의 AAC 부호화기가 128kbps의 MPEG-1 Layer 3(MP3) 부호화기가 제공하는 음질에 필적하는 음질을 제공하는 것으로 나타났다. 따라서, 동등한 음질 레벨에서 AAC는 MP3보다 약 1/3배 정도 큰 압축률을 제공한다.

AAC는 블록 기반의 가변 율 부호화(variable rate coding) 알고리즘이다. 즉, AAC 인코더는 1024 샘플의 입력신호를 읽어 들여 입력 데이터 블록을 가변 수의 비트로 압축, 표현한다. 인코더에서 율 제어(rate control)를 사용하여 출력 비트율이 일정한 전송율을 갖는 통신 채널에 요구되는 미리 결정한 값으로 평균된다. AAC로 압축된 각각의 비트 블록은 raw data block이라고 하며, 이것은 단독으로, 즉 이전 비트스트림 블록에 대한 정보를 사용하지 않고 디코딩될 수 있는 좋은 성질을 가

지고 있다. 이와 같은 성질은 패킷 통신 채널에 있어서 이상적이다. 즉, 만약 패킷 payload가 단일 raw data block이면, packet framing이 인코더와 디코더의 동기를 쉽게 할 수 있고, 또 가장 중요한 것으로 하나의 패킷의 손실이 인접 패킷의 디코딩을 방해하지 않는다는 것이다.

RTP payload는 32 또는 64 비트 헤더, 손실된 AAC 프레임의 복원하는데 필요한 정보를 포함하고 있는 가변 수의 RepairData 및 가변 수의 AAC 프레임으로 구성된다. 헤더는 근본적으로 디코더가 원래 신호를 복원할 수 있게 현재 및 과거 패킷의 priority를 규정한 priority quantizer(PQ) 벡터를 포함한다. X 비트는 헤더가 12 또는 28개의 PQ를 포함하는지를 나타낸다. REPAIRLEN는 RepairData를 포함하는 word(32비트)의 수를 규정한다. RepairData가 없으면 REPAIRLEN는 0으로 설정되어야 한다. 모든 REPAIRDATA 또는 AAC FRAME은 sequence number인 (R)SEQ와 길이 규정자인 (R)LEN를 선두부에 포함한다. 분할된 AAC 프레임의 경우에는 프레임의 분할이 byte-aligned되지 않을 수 있기 때문에 UBITS 필드가 최종 바이트에서 사용되지 않은 비트 수를 규정한다. 해당 프레임이 분할되지 않으면 UBITS는 0으로 설정하여야 한다. 그림 3에 RTP를 이용한 MPEG-2 AAC payload packet의 포맷을 나타낸다[3].

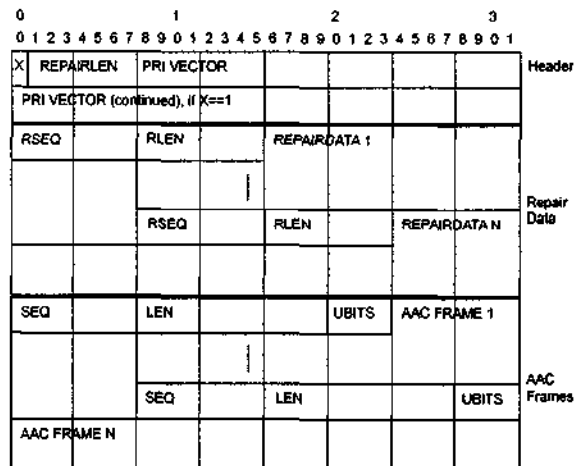


그림 3. MPEG-2 AAC RTP payload packet

3.2 MPEG-4 오디오

MPEG-4 오디오는 많은 서로 다른 형태의 오디오 부호화 도구(tool)들을 통합한 새로운 형태의 오디오 부호화 표준이다[4]. 또한, 이것은 합성음을 표현하는 메커니즘을 지원한다. 저 오버헤드(low-overhead)의 MPEG-4 오디오 전송을 위한 다중화인 LATM(Low-overhead Audio Transport Multiplex)은 MPEG-4 오디오 도구들에 의하여 압축된 또는 표현된 오디오 데이터 시퀀스를 상대적으로 낮은 오버헤드로 관리한다[5]. 따라서, 오디오만의 응용에서는, LATM 기반의 MPEG-4 오디오 비트스트림을 MPEG-4 시스템을 사용하지 않고 RTP 패킷으로 직접 매핑하는 것이 바람직하다.

LATM은 하나 또는 그 이상의 오디오 프레임들을 포함하는 audioMuxElements의 시퀀스로 구성된다. 하나의 audioMuxElement 또는 다수의 audioMuxElement의 일부

는 그림 4와 같이 어떠한 audioMuxElement 신덱스 요소의 제거 없이 RTP payload 위에 직접 매핑되어야 한다. 각각의 audioMuxElement의 첫번째 바이트는 RTP 패킷의 첫번째 payload에 위치해야만 하며, 각 헤더 필드의 정의는 다음과 같다[6].

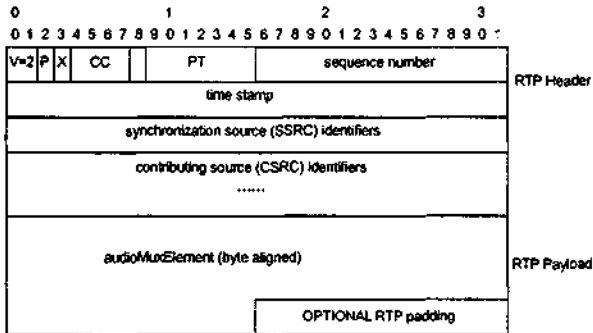


그림 4. MPEG-4 오디오 스트림 RTP 패킷

- PT(Payload Type) : MPEG-4 오디오 RTP payload 포맷을 규정하기 위한 특정 PT를 할당하는 것이 바람직하다.
- M(Marker) 비트 : 이 비트는 audioMuxElement의 경계를 나타낸다. RTP 패킷이 하나의 완전한 audioMuxElement 또는 audioMuxElement의 최종 fragment를 포함하는 것을 표시하기 위하여 M 비트를 1로 설정한다.
- Timestamp : 시간표는 합성(composition) 시간 또는 compositor가 없는 복호화에 있어서 표현(presentation) 시간을 가르키며, 보안상의 이유로 난수(random value)로 시작하는 것을 권고한다. 시간표의 분해능은 기본 값은 90kHz로 설정된다.
- Sequence Number : 각각의 RTP 데이터 패킷이 전송될 때마다 1씩 증가한다. 보안상의 이유로 초기값은 난수로 시작한다.
- SSRC, CC와 CSRC 필드는 RFC 1889[7]에서 기술한 대로 사용된다.

RTP 패킷당 하나의 audioMuxElement를 두는 것이 바람직하다. audioMuxElement의 크기는 결과적인 RTP 패킷이 path-MTU보다 크지 않도록 조절된다. 만약 이것이 불가능하면 다음의 원칙에 따라 audioMuxElement를 여러 개의 패킷에 걸쳐 분할할 수도 있다.

- ① payload 요소로 구성되는 "payloadMux"를 여러 개의 RTP 패킷으로 분할하여 하나의 RTP 패킷이 한 개 이상의 payload 요소로 구성되도록 할 수 있다. 하나의 payload 요소는 분할되지 않도록 하는 것이 좋다.
- ② 만약 audioMuxElement가 StreamMuxConfig를 포함하고 있으면 첫번째 payload 요소를 포함하는 RTP 패킷이 StreamMuxConfig를 포함해야만 한다.

4. 기술개발을 위한 고려사항

현재 국내의 인터넷 방송이나 웹캐스팅(Webcasting)에 사용되고 있는 AV 스트리밍 기술은 외국에서 개발된 시스템이 주로 사용되고 있는 실정이며, 대부분의

시스템은 저작권을 가지고 있으며(proprietary), 기술적인 스펙은 주로 일반에게 공개되지 않고 있다.

그러나, 인터넷 환경에는 전통적인 방송에서의 같은 더 이상 단일 부호화 시스템에 대한 요구가 존재하지 않으며, 많은 종류의 오디오 코덱의 출현으로 서비스 제공자에 의하여 비용, 오디오 품질, 서버의 신뢰성, 실시간 처리 능력, 서비스 품질, scalability 등을 고려하여 그들 자신의 부호화 방법이 선택될 것이다.

따라서, 현재 새롭게 부각되고 있는 MPEG-2 AAC 및 MPEG-4 오디오를 이용한 스트리밍 분야에서는 우리 자체의 기술을 확보해야 할 필요성이 증대하고 있으며, 이 때 고려하여 할 사항을 다음과 같이 제안한다.

- ① MPEG-2 AAC/MPEG-4 오디오 압축 기술 및 압축 기술에 적합한 실시간 scalable 스트리밍 기술의 개발
- ② MPEG-2 AAC에 적합한 RTP payload 포맷을 이용한 스트리밍 기술 개발 및 RTCP(Realtime Control Protocol)를 이용한 전송 모니터링 기술 개발
- ③ RTP/RTCP를 이용한 MPEG-4 오디오 payload 포맷 제안 및 스트리밍 기술 개발 : MPEG-4 시스템 기반, ATM 기반, 그리고 Flex-Multiplexing 기반의 각각의 payload 포맷 및 스트리밍 기술

5. 결론

본 논문에서는 인터넷 오디오 방송을 위한 오디오 스트리밍 기술 동향을 분석하고 현재 IETF에서 논의되고 있는 MPEG-2 AAC 및 MPEG-4 오디오를 인터넷을 통하여 스트리밍하기 위한 payload 포맷을 기술하고, 또한 기술개발을 위한 고려사항을 제안하였다.

MPEG-2 AAC 및 scalable 기능을 이용한 MPEG-4 오디오 스트리밍 기술이 앞으로 인터넷 오디오 방송 서비스를 위한 핵심기술로 부각될 것이기 때문에, 향후 이를 위한 payload 포맷 개발, 실시간 전송을 위한 RTP/RTCP 기술 및 오디오 데이터의 저작권 보호를 위한 IPMP 기술 등이 개발되어야 할 것이다.

감사의 글

본 연구는 정보통신부의 "대화형 디지털 방송기술 연구" 사업의 일환으로 수행된 연구결과입니다.

참고문헌

- [1] Gerhard Stoll, "Streaming-audio@Internet: Perspectives for the broadcasters", AES 17th intl' conference on high quality audio coding, Italy, 1999.
- [2] ISO/IEC 13818-7 Advanced Audio Coding (AAC), 1997.
- [3] M. Kretschmer et. al., "RTP payload format for MPEG-2 AAC streams", draqft-ietf-avt-rtp-mpeg2aac-00.txt", Jun. 1999.
- [4] ISO/IEC 14496-3: 1999, "Information technology - Coding of audio-visual objects - Part 3: Audio, Dec. 1999.
- [5] ISO/IEC 14496-3: 1999/FDAM1: 2000, Dec. 1999.
- [6] Y. et. al., "RTP payload format for MPEG-4 Audio/Visual streams", draqft-ietf-avt-rtp-mpeg4-es-01.txt", May. 2000.
- [7] Schulzrinne, Casner, Frederick, Jacobson, "RTP: A Transport Protocol for Real Time Applications", RFC 1889, IETF, Jan. 1996.